AFIT/GCA/LAC/96S-6

AN EVALUATION OF THE FACTORS USED TO PREDICT WRITING ABILITY AT THE AIR FORCE INSTITUTE OF TECHNOLOGY

THESIS

Darrin E. Farr, Captain, USAF
AFIT/GCA/LAC/96S-6

The views expressed in this thesis are those of the author
and do not reflect the official policy or position of the
Department of Defense or the U.S. Government.

AFIT/GCA/LAC/96S-6

AN EVALUATION OF THE FACTORS USED TO PREDICT WRITING ABILITY AT THE AIR FORCE INSTITUTE OF TECHNOLOGY

THESIS

Presented to the Faculty of the Graduate School of

Logistics and Acquisition Management of the

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science in Cost Analysis

Darrin E. Farr, B.S.

Captain, USAF

September 1996

Acknowledgments

In the course of this research, I have been able to overcome many problems, and I would like to thank some of the individuals who helped me overcome these difficulties. Dr. David Vaughan, my advisor, provided a great deal of encouragement and direction during this effort. Thank you for the many expedient reviews. Lt Col Stephen Atkins, my reader, and Lt Col James Van Scotter provided the much needed quantitative assistance. Thanks for the enthusiasm and direction.

The next group of individuals were extremely helpful during the data collection portion of this effort. Major Cheryl Lee (AFIT/RR) and her people, SrA Matthew Joseph (AFIT/SC), and Mr. Richard Gainor (AFIT/CI) were extremely courteous, expedient and professional in the performance of their duties. Don Powers and Robb Durso from the Educational Testing Service also provided a great deal of information. Thank you all very much.

A hearty thanks go out to all of my classmates. Rarely does one have the pleasure of working with such a fine group of people.

Finally I would like to thank my wife, Amy, for her love and support through this time. I also want to thank my son, Matthew, for the many (sometimes too many) very enjoyable study breaks.

Darrin E. Farr

Table of Contents

Page

Page

List of Figures

Figure Page

List of Tables

Abstract

A study of 574 students at the Air Force Institute of Technology compared performance, education, and experience factors--the later two as stated by the students themselves--to a locally developed estimate of true writing ability (WGPA). This exploratory research was additionally intended to assess the effectiveness of AFIT's current writing student skill diagnostic and instructional system. Direct (essay evaluation) and indirect (objective test) evaluations of AFIT student writing ability were analyzed for their predictive impact. The statistical analysis procedures used in this study included the factor analysis of a survey, ANOVA, the adjustment of multiple correlations due to measurement error and range attenuation, and the performance of a regression analysis using the raw data and the adjusted correlation matrix. The results of this study indicate AFIT's direct evaluation portion (essay examination) is useful for determining writing ability; the indirect portion (objective test) did not significantly contribute to the model. Due to the combination of independent variables chosen for the predictive model, the study was unable to identify the immediate benefits of the written communications review course on AFIT performance.

AN EVALUATION OF THE FACTORS USED TO PREDICT WRITING ABILITY AT THE AIR FORCE INSTITUTE OF TECHNOLOGY

I. Introduction

Educational institutions, like the Air Force Institute of Technology (AFIT), attempt to foster written communications skills throughout their educational programs. By more accurately identifying students who may require additional instruction, the Air Force Institute of Technology will be better able to prepare students for the writing assignments they will receive in graduate school.

The focus of this research is to determine whether performance, education, and experience factors can be correlated with writing ability, and then to use these correlations to support an initial assessment of the current Air Force Institute of Technology (AFIT) written communications review course. A study of 574 students at the Air Force Institute of Technology compared performance, education, and experience factors--the later two as stated by the students themselves--to a locally developed estimate of true writing ability (WGPA). This exploratory research was additionally intended to assess the effectiveness of the current writing student skill diagnostic and instructional system.

English Proficiency Examination (EPE)

Simultaneously with the survey, the EPE is administered. The EPE is a locally-designed test which is given to all incoming students to assess their English proficiency. The EPE is a two-part examination which directly and indirectly evaluates the student's writing ability. The direct or essay portion is graded as a whole considering the essay's organization, completeness of thought, clarity of style, factual or illustrative support, and correctness of grammar, punctuation and

mechanics. The indirect or objective portion of the EPE contains twenty-five multiple choice questions designed to test the student's ability to identify incorrect grammatical constructions, and given a sentence, select the proper words or phrases to correctly complete the sentence. Students have one hour to complete the essay and 30 minutes to complete the objective portion. The scores received on the two portions of the EPE are used to identify students who may need additional writing instruction.

The additional writing instruction is a 4-week course providing 20 hours of instruction in the basic aspects of written expression, including grammar, mechanics, usage, style, logic, and organization (1993-1995 AFIT Catalog: 168).

Factors

Performance factors include undergraduate grade point average (UGPA), Graduate Record Examination Verbal and Quantitative sections (GRE-V and GRE-Q), and Graduate Management Admission Test (GMAT) scores. These factors are available through the AFIT Registrar's Office. Education factors include the student's undergraduate major, the number of English composition courses completed, and length of time out of school. Experience factors include the amount of writing accomplished in previous job assignments, on-the-job writing instruction, and the individual's assessment of his or her writing skills. The education and experience factors were gathered from AFIT students using a 20-question survey asking students to describe their education and experience. The survey is administered during student orientation, and results have been obtained for an eight-year period for a total about 1000 students.

Writing Grade Point Average (WGPA)

An additional criterion measure was used in this research to approximate the student's true writing ability. This criterion is the student's writing grade point average, or WGPA. The

WGPA construct is the student's grade point average from AFIT courses which are considered writing intensive by the faculty members. In a writing intensive course, a major portion of the grade is directly attributable to the student's writing ability as assessed by that particular instructor. Thirteen AFIT graduate courses have been identified as writing intensive. Since the relationship between writing ability predictors and WGPA might differ based on enrollment in the English review class, the effects of enrollment in the class were analyzed simultaneously with the predictors.

Research Questions

The research questions examined here are: Can performance, education or experience factors be correlated to writing ability at the graduate level? If so, does the EPE provide any additional information about the student's writing ability above what is known from AFIT records and the student survey? The final research question is: Is there significant value-added by requiring less able students to complete the 20-hour remedial writing course? To answer these questions, additional investigative questions must also be answered. What does research say are relatively reliable indicators of effective writing ability? What are the expected strengths of the correlations between these indicators and writing ability? Can effective writing ability be measured? What is the reliability of AFIT's current testing technique?

The following chapters discuss the related literature, methodology, analysis and results. The literature review focuses on assessing writing ability and the use of standardized test scores as predictors. The methodology chapter explains the different techniques used and the steps taken to analyze the data and draw conclusions. Chapter IV, the statistical output is presented; Chapter V discusses the conclusions, future research and study limitations.

II. Literature Review

In this chapter, research related to writing assessment is examined. The research reviewed deals with combining direct and indirect measures to assess writing ability, the use of holistic essay grading, standardized test usage, and reliability.

Essay Evaluation

Essay evaluation has been used to assess writing ability since the 1880s. Researchers have long attempted to quantifiably measure writing ability; however, this has proven to be a difficult task. Early studies involved the direct evaluation of a writing sample. Intuitively, essay evaluation would seem to be the best measure of writing ability; however, essay evaluation efforts were complicated by unreliability (Huddleston, 1954; Follman and Anderson, 1967). The scores received on an essay depended greatly on the bias of the grader and on other peripheral factors such as essay length, subject, and handwriting quality (Breland et al, 1987). One early study showed dramatic variations in score based solely on penmanship (Sheppard, 1929). Because of the relatively low reliability associated with essay assessment, other methods of writing skill evaluation were developed (Bradley-Johnson and Lesiak, 1981). Research has shown writing skill is best assessed by using both direct and indirect assessments in the form of essays and multiple choice examinations (Breland et al, 1987).

The most prevalent indirect method of writing evaluation is multiple choice examination; however, many educators and employers believe there is no substitute for an actual writing sample. As Diederich observes:

> As a test of writing ability, no test is as convincing to teachers of English, to teachers in other departments, to prospective employers, and to the public as actual samples of each student's writing, especially if the writing is done under test conditions in which one can be sure that each sample is the student's own unaided work. (1974)

Direct and Indirect Assessments

Research indicates that good predictions of writing ability may be made by utilizing direct and indirect assessments (Breland et al, 1987). The relative unreliability of the essay evaluation can be partially compensated by the increased reliability of the indirect assessment technique, while the direct assessment adds credibility to the multiple choice examination. In Breland's 1987 study, direct and indirect assessments were made of the work of college freshmen. The students completed six different essays which were read by multiple readers. The students also completed the English Composition Achievement Test (ECT), SAT-Verbal or the Test of Standard Written English (TSWE). The combination of one reader for the essay and either of the standardized tests produced an average multiple correlation with writing performance of approximately .70. In fact, when essay and non-essay assessments are combined, the number of different essay readers becomes much less important (Breland et al, 1987: 32). The writing performance criterion is Breland's measurement of the student's writing ability. This criterion is comprised of other essay assessments, instructors' judgments of students' writing skill, and grades assigned in a college freshman English composition course.

Reliability

As mentioned before, essay examinations are often plagued with the problem of unreliability; however, means are available to increase their reliability, and results can be adjusted to control the affects of unreliability. The term reliability indicates the extent to which essay and test scores are consistent. Spandel (1980) presents a simple scenario in which the reliability among raters with unknown reliability is a problem. Suppose a student writes an essay which is assessed by two different raters, and each rater grades the essay differently. How would one assess the student's writing proficiency? Rater reliability would provide some measure of the accuracy of the rater's assessment.

Three factors are important to developing a reliable test. The writing skills measured must be clearly defined and understood by both writer and assessor, the test must be clearly linked to the skill being measured, and the raters must be carefully trained in the appropriate evaluation standards (Spandel, 1980: 10).

Holistic Grading

Recent evaluation efforts have focused on the holistic approach to essay scoring. In holistic grading essays scores are based on the overall impression of the essay. Factors like organization, completeness of thought, clarity of style, factual or illustrative support, and correctness of grammar, punctuation and mechanics affect the rater's response; however, these characteristics are not individually measured. Holistic scoring is a widely used technique valid for ranking students (Breland 1983: 20). It is rapid and efficient; however, ensuring reliability when grading essays is a major concern. When holistic scoring is used, consistency is very important among scores assigned by the rater and among different raters.

Educational Testing Service (ETS) Training

One way to increase the reliability is to train the raters. The principal purpose for using trained raters is to minimize the effects of individual biases by helping the rater conform to specified scoring standards (Spandel, 1980: 19). The Educational Testing Service (ETS) has conducted highly standardized readings of essays for several years. In the ETS system, essay readers are brought together in a large group consisting of from 100 to 800 readers, depending on the number of essays to be read. The readers are grouped at tables, usually nine readers to a table, with an experienced reader serving as table leader. The table leader supervises the readings of the other readers at the table. There are group leaders as well, and the entire reading is under the supervision of the chief reader and an assistant chief reader.

During the opening session, readers read sample essays selected from the essays to be read. These sample essays have been previously read by the table leaders, group leaders, and chief reader, and they have decided the scores the essays should receive, using the standard ETS holistic scoring scale. Scores of 4, 5, and 6 are upper-half scores, with 6 being the highest score. Scores 1, 2, and 3 are lower-half scores, with 1 being the lowest score. The characteristics of each of these six scores are discussed in the group, and the group practices scoring the sample essays. The results are discussed by the table leaders; any readers whose evaluation diverges noticeably from the established scores are counseled and urged to reconsider their responses to the essays.

After a two-hour training session, the readers begin to read live, unscored essays. The table leaders circulate among the readers at their tables, reading scored essays at random to ensure consistency in grading, and selected scored essays are distributed to group leaders to ensure scoring consistency as well. Periodically throughout the reading--which may run for two, three, or four days, sample essays are read, scored, and discussed by the entire group to make sure that readers remain consistent in their scoring as the reading progresses. As a result of thorough training and monitoring of the essay evaluation process, essay scores are remarkably consistent and appropriate.

The primary evaluator of the AFIT student essays is an experienced Educational Testing Service (ETS) reader. He has participated in over ten ETS essay readings since 1987 and has utilized his ETS training in evaluating AFIT student essays.

Study Expectations

This study is similar to Breland's 1987 study, which evaluated the writing ability of college freshman based on standardized test scores as well as essay evaluations. Although Breland used the ECT, SAT-V and TSWE scores for his indirect measure of writing ability, the analogous measures used in this study are the GRE and GMAT scores and the EPE objective portion. This

research correlates direct and indirect measurements of writing ability for graduate students, and therefore the graduate level standardized tests were used. Breland found that an essay read once by a single reader and the standardized tests correlated with writing performance (with a correlation coefficient of about .56 and .64 respectively). However, their combined correlation becomes about .70 (Breland et all, 1987:59). Based on this observation, analogous results for graduate students are expected from this study.

Standardized Tests

Standardized tests constitute a major part of the performance factors considered in this study. In this study, the writing grade point average (WGPA) and EPE test results are correlated with GRE-Verbal, GMAT and undergraduate grade point average to discover the extent to which the standardized tests can be used to predict writing performance. Of interest to this study is the observed correlations between standardized test scores and graduate grade point average (GGPA) from other studies.

Nilsson's study investigated the relationship between GGPA and the GRE and GMAT tests respectively. This study of 60 master's degree students, 30 of whom had taken each of the tests, revealed a stronger relationship between the GRE and GGPA. The Pearson Moment Coefficient was .449 between GRE and GGPA and .231 between GMAT and GGPA. The GRE students in the study were enrolled in Masters programs in English, education, gerontology, history, psychology, sociology, rural and small town planning or music. The GMAT students were enrolled in business administration, professional accounting or public administration programs. Nilsson recognizes as study limitations the small sample size and the diversity of educational programs. However, a previous AFIT study which involved over 2000 students has shown overall graduate grade point average (GGPA) is moderately correlated with GRE-V, GRE-Q, and GMAT scores. The correlation coefficients for each respectively were .163, .351 and .440 (Van Scotter, 1983). Since

Van Scotter's work specifically pertains to AFIT and the sample size was considerably larger, this study uses GMAT or an estimated GMAT score to predict writing ability.

Reliability

The reliability of the essay examination in this study was controlled by several study-specific factors. In accordance with Spandel's (1980) three factors mentioned earlier, the students were given, as part of the instructions in the EPE, a focused essay topic and the specific characteristics considered in the scoring process. The topic addresses a current military or defense-related issue so the students, who are all military officers or government civilians working for the Department of Defense, will have some familiarity with the essay topic. The students are given extra incentive to excel in this task because a passing score allows the students to test out of a review course. A very limited number of raters (typically one) trained in essay scoring by the Educational Testing Service (ETS) evaluated subject essays. This training positively impacts the reliability of the results and helps to reduce the grade variation among year groups by increasing rater consistency.

Summary

This literature review was intended to validate the approach used in this study. Similar to Breland's 1987 study, in which his criterion for writing performance consisted of grades from writing courses, the measure of writing ability in this study is the average grade received in writing intensive courses completed at AFIT. Breland used indirect and direct measures and determined their correlations with writing performance. This study uses the indirect measures, AFIT records

and an objective test, and a direct measure, the essay, to determine correlations. Spandel (1980) served to validate the methods employed to increase the reliability of this study.

III. Methodology

This chapter explains how the data were collected and the analysis techniques used in this study. The data consists of three main groups: performance measures, which were gathered from AFIT records; English Proficiency Examination results; and the accompanying survey data. Parametric statistical techniques were used to analyze the data. The specific techniques used consisted of factor analysis, Pearson Moment for correlation, correlation adjustments, single factor analysis of variance (ANOVA), Levene test for homogeneity of variance, and multiple regression techniques.

Data Overview

The types of data used in this study relate to measures of the students' general academic aptitude (GRE and GMAT scores), their past academic performance (undergraduate GPA or UGPA), an approximation of their writing performance while attending AFIT (WGPA), a single measure of performance potential (EPE), and self-reported education and experience measures (survey). These data were grouped into three general measures: Performance, Education, and Experience. Performance measures included data from standardized testing, undergraduate grades, and results from a locally developed English Proficiency Examination (EPE). Educational factors include the student's assessment of their high school English preparation, number of English composition or literature courses completed in high school and college, a self-assessment of writing ability and others. Experience factors include the student's perception of the importance of writing skills in his or her career, amount of professional writing instruction received, amount of technical writing or report preparation done on the job, amount of writing instruction given and the amount of career related guidance received. Other data collected on the survey included the source of commission (ROTC, OTS or Air Force Academy) and the type of undergraduate degree

(Humanities/Social Sciences, Science/Technical, Mathematics, business or other). The database began with the information from over 1000 student surveys. Approximately 574 complete student records were used for this study. A complete student record contains undergraduate grade point average, either GRE or GMAT scores, writing grade point average, and survey information.

Performance Metrics

Most of the performance metrics are available through AFIT records. These records include GRE and GMAT scores, UGPA, and AFIT writing GPA (WGPA). The GRE and GMAT examinations are similar, but the GMAT focuses on students interested in business and management graduate programs. These tests measure the verbal, quantitative and analytical skills developed over time. This information is collected to assist graduate program selection by providing the institution with a measure associated with success in the first year of graduate study (1995-1996 GMAT guide). The data available for the GRE is broken out into scores for the Verbal and Quantitative portions of the exam. Only total scores were available for the GMAT. A previous AFIT study has shown overall graduate grade point average (GGPA) is moderately correlated with GRE-V, GRE-Q, and GMAT scores. The correlation coefficients for each respectively were .163, .351 and .440 (Van Scotter, 1983). Since GMAT displayed the strongest correlation with GGPA, it was chosen as the standard measure in this study for standardized test scores. The substitute used in this study was called the Graduate Management School Aptitude (GMSA). The GMSA is either the actual GMAT score (if available), or the weighted sum of GRE-V and GRE-Q. This relationship is based on the following regression equation:

$$GMSA = 70.393 + .533 * GRE\text{-}Q + .282 * GRE\text{-}V$$

The regression produced an adjusted r^2 value of .5221, and both beta weights and the model are significant at the $\alpha = .0001$ level (n=78).

The English Proficiency Examination (EPE) (Appendix A) is designed to test the grammar and writing skills of the incoming students during AFIT's orientation phase. The EPE consists of an objective test and an essay. The primary purpose of the writing assessment at AFIT is to identify students who would benefit the most from a review of the fundamentals of written communication. For the past eight years, student survey information and preliminary English Proficiency Examination scores have been recorded. The multiple choice portion evaluates the student's ability to recognize basic stylistic and structural conventions of standard written English, and the essay portion evaluates the ability to clearly and concisely organize thoughts. The students are given 30 minutes to complete the 25-question multiple choice grammar examination and 1 hour to compose an essay. The multiple choice examinations are scored by totaling the number of correct answers. The essays are given a score ranging from 1 to 10, 10 being the best, based on the grader's holistic impression of the essay. If the essay is sound, the argument is usually clearly organized around a hierarchy of goal and subgoals (Freedman et al, 1987). A "10" essay is clear, well organized, complete, supported, and uses correct grammar, punctuation and mechanics.

Unfortunately, the only data recorded pertaining to the EPE were the objective test and essay scores. If the information had been available, the objective test answers could have been factor-analyzed to determine the number and composition of the subscales which were actually being tested and the reliability of the test would be determined empirically. Likewise, the essays have not been graded using multiple trained readers and the actual essays have not been saved over the years so there is no way to determine the grades' actual reliability.

The survey (Appendix B) was intended to provide a clearer understanding of the student's education, background, and training in the area of written communications skills. The information helps AFIT design writing instruction to better suit the student's needs. The student survey given each year captured the education and experience items related to this study.

Education and Experience Factors

These factors were extracted from the student survey using factor analysis to form composites for both education and experience. The underlying assumption was that more education and more recent writing experience were expected to be associated with better writing skills. The survey questions designed to capture education and experience factors are as follows. The possible answers are in parentheses after the question with specific responses separated by a comma.

Education Factors

Nine questions were linked to individual education aspects:

1. What is the source of your commission in the armed forces? (ROTC, USAF Academy, OTS, Other, Not applicable - civilian/foreign military)

3. What was the area of concentration of your undergraduate degree? (Humanities, Social science, Business, Engineering, Mathematics/Science, Other)

5. How would you assess your high school/pre-college level English preparation (English classes)? (Excellent, Good, Average, Fair, Poor)

6. What size college or university did you attend? (Very large (over 10,000 students), Large (5,000-10,000 students), Medium (2,000-4,999 students), Small (Under 2,000 students))

7. In how many undergraduate (college/university) writing courses (English) did you enroll? (0, 1, 2, 3, 4, More than 4)

8. In how many undergraduate literature courses did you enroll? (0, 1, 2, 3, 4, More than 4)

9. In how many high school English classes did you enroll? (0, 1, 2, 3, 4, More than 4)

10. How long has it been since you completed your most recent undergraduate (or graduate) degree? (1 year or less, Less than 2 years but more than 1 year, Less than 3 years but more than 2 years, Less than 4 years but more than 3 years, Less than 5 years but more than 4 years, Less than 6 years but more than 5 years, 6 or more years)

15. What is your estimation of your writing ability? (Well above average, Above average, Average, Below average, Well below average)

Experience Factors

Eight questions were linked to professional experience:

2. What is your rank? (2nd Lieutenant, 1st Lieutenant, Captain, Major, Lt Colonel, International Military, Civilian)

4. How would you describe the value of good writing skills in your career? (Essential, Very important, Moderately important, Somewhat important, Not important)

11. In your most recent assignment, how much of your writing did you do on a computer word processing? (All, Most, Some, Very Little, None)

12. In general, how would you describe the writing instruction you have received while working for the government or military? (Very Good, More than adequate, Adequate, Less than adequate, Poor)

13. Assuming that technical writing includes proposals, mechanism descriptions, process descriptions, or sets of instructions, would you say that you have done some technical writing on the job? (Yes-a significant amount, Yes-some, Yes-a little, No)

16. Have you ever provided writing instruction to others in your military or civilian profession? (Yes- formally in a class, Yes- informally in a working environment -- quite a bit, Yes- informally in a working environment -- some, Yes- informally in a working environment - a little, No)

17. How much has writing or report preparation been a part of your military or civilian career? (A crucial component: over 75% of time spent writing, A major component: 51-75% of time, A moderate component: 26-50 % of time, A minor component: 11-25% of time, A negligible component: O-10% of time)

18. How much formal career-related writing instruction have you received? (A significant amount, A moderate amount, A small amount, Very little, None)

Questions 19 and 20, which identified program specialties, were not considered in this study because the program designations have changed over the years and quality of this data within the data base is suspect due to scanning difficulties incurred when the data was converted to electronic format.

The source of commission and type of undergraduate degree obtained were the two qualitative categorical factors considered. The expectations were that Air Force Academy (AFA) graduates may possess better writing skills than ROTC and OTS graduates due to more stringent entrance requirements. The undergraduate degree types were divided into technical and non-technical degrees with mathematics, science and engineering constituting technical degrees while business, humanities and social sciences comprise the non-technical degree types.

Statistical Analysis

Histograms were constructed of variables to visually identify variables which might be not be normally distributed. Parametric statistical techniques were used to analyze the data for this study because all of the variables were approximately normally distributed.

Factor Analysis. The factor analysis used in this study was able to group the survey questionnaire into two distinct sub-scales for education and experience factors. The goal of factor analysis is to identify some combination of items which can be combined into a logical composite. This composite then serves as the predictor (so as to avoid using the subordinate items as single severely intercorrelated predictors of unknown reliability).

In the factor analysis, the method of factor extraction chosen was principal axis factoring (PAF). PAF uses squared multiple correlations (SMCs) to estimate the total amount of common variance in the variables potentially accounted for by consideration of factor levels and loadings. Specifically, the SMCs (when used as initial estimates of communality) equal the squared multiple correlations that result from trying to predict outcomes on one variable from all the other variables in the analysis. This PAF approach is thought to reduce idiosyncratic capitalization on chance arguably associated with using 1.0 as a communality estimate. These SMCs are estimated for each variable one at a time. They replace the 100% communalities in the diagonal of the observed correlation matrix. Factor analysis then iteratively adjusts factor loadings trying to find the optimal set of relations among items that will yield a "reproduced" correlation matrix as similar as possible to the observed correlation matrix (altered with SMCs on its diagonal). Visual inspection of eigenvalues plotted against factors (the so-called Scree plot) indicated a two factor solution. Specifying two factors via SPSS direct oblimin (an oblique rotation seeking simple structure for correlated factors) yielded two sub-tests. These were the Education and Experience factors described later.

Correlation Coefficients. The Pearson product moment coefficient of correlation, r, is a numerical measure of linearity between two variables. If two variables are positively correlated, when the first variable is increased, the second will also increase. Conversely, negatively correlated variables move in opposite directions from their respective means. The lower case 'r' is used to represent correlation. The absolute value of r is between 0 and 1. The larger the value, the more strongly the variables are said to be correlated. Computing the correlation coefficient for each variable as it relates to the other variables produces a correlation matrix. The formula used to compute r is as follows:

$$r = SS_{xy} / (SS_{XX}SS_{YY})^{.5}$$

where:

$$SS_{xy} = \sum XY - [(\sum X)(\sum Y)/n]$$

$$SS_{xx} = \sum X^2 - [(\sum X^2)/n]$$

$$SS_{yy} = \sum Y^2 - [(\sum Y^2)/n]$$

The correlations are used as an initial indication of variable interaction. In this study, they were adjusted for measurement error and range attenuation and the resulting correlation matrix was used as part of in the multiple regression analysis discussed later (McClave, 1991:665).

Adjustments to Correlation Coefficients. Because one of the goals of this study is to identify factors correlated with writing skills, reliability is an important component of this research. If measurement error or unreliability is substantial, the observed correlations between variables will be lower than the true correlations between the variables (Nunnally, 1973: 238). According to

Nunnally (1973), statistical corrections using reliability coefficients should be made to account for measurement error. Using the methods prescribed by Nunnally, it is possible to estimate the extent to which correlations between variables are attenuated by measurement error.

Nunnally (1973) shows the effects of measurement error on reliability. Reliability and measurement error go hand in hand. Measurement error decreases the reliability of a criterion and it is essential to correct for projected measurement error to produce an accurate estimation. If a measure's reliability is known, then the distortion present in the measure can be corrected. Formula 1 performs this correction:

$$r_{xy_c} = r_{xy}(\Delta r_{xx} * \Delta r_{yy})^{.5} / (r_{xx} * r_{yy})^{.5} \quad (1)$$

where r_{xy_c} is the corrected correlation, r_{xy} is the original observed correlation, r_{xx} is the reliability of X, r_{yy} is the reliability of Y, Δr_{xx} is the changed reliability of X, and Δr_{yy} is the changed reliability of Y. Schmidt and Hunter (1996) caution against assuming perfect reliability in the dependent variable of a regression analysis. This assumption may cause some beta weights to appear significant (although typically error variance in the dependent variable reduces the power or sensitivity of the analysis). The reliability in the dependent variable should be further investigated; however, it was considered beyond the scope of this thesis.

Range Restriction. Another factor affecting reliability is range attenuation. Range attenuation occurs in any selection process in which people are chosen from an underlying population based on specific criteria. The statistics pertaining to the population are seldom those of the selected sample. Since the sample mean and standard deviation are different from those of the overall population, the reliability coefficient will also be distorted.

> Unfortunately, this type of selection affects not only the absolute size of the validity coefficients but also their relative size, so that the test which is really most valid as applied to the general run of applicants may appear to be one of the less valid in a group resulting from high standards of pre-selection. The reduction in the validity of a test within a selected group becomes greater the more closely the test correlates with the basis of selection. (Pedhazur, 1982: 171)

In this study, the attenuated range of GRE, GMAT and undergraduate UGPA negatively impacts the correlation with EPE scores and WGPA. To obtain the adjusted correlation between variables, correction formula 2 is:

$$R_{12} = \{r_{12} * S_1/s_1)\}/(1 - r_{12}^2 + r_{12}^2 * S_1^2/ s_1^2)^{.5} \qquad (2)$$

where R_{12} is the corrected correlation between variables 1 and 2, r_{12} is the uncorrected correlation, S_1 and s_1 are the standard deviations of the unrestricted population and the restricted sample respectively (Pedhazur, 1982 : 173). This formula makes apparent that if $S_1 = s_1$, then $R_{12} = r_{12}/(1)^{.5} = r_{12}$, and that R_{12} increases as S_1 increases with respect to s_1.

Analysis of Variance. An ANOVA is a statistical technique designed to determine if significant differences exist among the means of normally distributed populations with equal variances. The null hypothesis in an ANOVA is that all of the population means being tested are equal, and the alternative is that at least one is different. The Levene test was used to ensure the

samples being compared had approximately equal variances (McClave, 1991 pg. 462). The Levene Test for homogeneity of variance tests whether the mean of the absolute deviations for one group is significantly different than that of another. The test uses the same structure as a basic t-test (the difference is a modified test statistic). If the significance level (i.e., risk of Type I error) of the test statistic is less than .05, the groups were not considered homogeneous with regards to variance (Neter et al, 1990: 112).

Multiple Regression - Standard. Multiple regression refers to the development of a probabilistic model in which two or more independent variables are used to predict a value for a dependent variable. The general form of a multiple regression equation is:

$$y = b_0 + \sum b_i x_i + \varepsilon \quad (3)$$

In this equation, b_0 represents the y intercept and b_I is a weighting factor assigned to each of the independent variables which minimizes the sum of the squared error terms derived from the actual data points associated with deviations from the predicted line minus their predicted points. The final term, ε, is the normally distributed random error component which has a mean of zero. This study uses multiple regression models to determine the capability of the performance, education and experience factors (coupled with EPE scores) to predict WGPA.

Multiple Regression - Correlation Matrix. Due to the impact of measurement error and range attenuation on the raw data, the adjusted correlation matrix was entered into SAS and the regression analysis was performed once again. Using this procedure, according to Pedhazur (1982:80), each of the regression beta weights, as shown above, can be derived using the correlations between each of the independent variables and the dependent variable. Empirically this is stated as:

$$b = \mathbf{R}^{-1}\mathbf{r} \quad (4)$$

where: b is the standardized variable coefficient. $\mathbf{R}^{-1}$ (in matrix notation) is the inverse of the correlation matrix, and **r** is the column vector for dependent - independent variable correlations. The unstandardized coefficients are calculated with the formula:

$$b_i = b_i(s_y/s_i) \quad (5)$$

where: b_i is the unstandardized beta weight for independent variable i, b_i is the previously calculated standardized weight, and s_y and s_i are the standard deviations for the dependent and independent variables respectively.

This chapter has detailed the methodology employed in the analysis of these data. In the next chapter, the analysis is performed and the results are presented.

IV. Analysis and Results

In this chapter, the analysis steps are identified and the resulting data are presented. The data analysis performed for this research was accomplished in the Statistical Package for the Social Sciences (SPSS™) and SAS™. Once the data were standardized within the database, the analysis process began. A factor analysis was performed on the survey data in an attempt to combine some manifest indicators into a measure which could more reliably relate a facet of writing experience to our criterion, WGPA, than the individual survey questions.

The distributions were then checked for normality. A histogram was constructed for each of the performance, education and experience factors. The students were then sorted by commissioning source and degree type and those groups were tested for significant differences among the essay and objective test scores using ANOVA and the Levene test.

A multiple correlation was performed to determine which predictors had statistically significant relations with AFIT's measures of writing ability. Once the correlated factors were found, scatter plots were drawn to spot evidence of a non-linear relationship. Correlation coefficients were calculated to analyze the relations between the criterion and the various predictors. These new correlation coefficients were then adjusted for range attenuation and estimated measurement error. Finally, a regression based on the adjusted correlation matrix was performed to investigate the possibility of a regression discontinuity (indication of an education effect for AFIT's remedial writing class).

Factor Analysis

Factor analysis is designed to group interrelated survey items into subscales which are assigned items measuring essentially the same construct. For example, several of the survey questions are designed to measure educational preparation. One would expect these items to be correlated with one another, and, therefore, educational preparation might be measured more

reliably if these items are combined into a single subscale (i.e., sub-test). An example of survey items which may be correlated would be the assessment of high school and pre-college level English preparation, the number of English, writing and literature courses taken, the amount of on-the-job writing assistance provided to others, and the student's assessment of their own writing ability. When the factor analysis was performed, as explained in Chapter III, two distinct factors were evident. The first factor included items 5, 7, 8, 9, and 15 (Coefficient alpha =.5496, n=1087). These items relate to educational experiences and individual assessment of writing skills; therefore, these items were called the Education (ED) factor. The second grouping of factors included items 4, 12, 13, 16, 17, and 18 (Coefficient alpha =.7009, n=1098). These items represent professional experience and the variable used to represent these items is Experience (EXP). The reported alpha values estimate the reliability of each factor. See Appendix C for the Item-total Statistics. Although the oblique rotation technique was used to produce a more conservative estimate of the factors' interdependence, the Factor Plot in Rotated Factor Space (Appendix C) shows these factors are very nearly independent (i.e., orthogonal).

Distribution Analysis

The histograms yielded no great surprises. The data appeared approximately normally distributed; however, range attenuation was evident in the standardized test scores and undergraduate grade point averages. The range attenuation results from AFIT academic selection processes. Students must meet certain academic standards to qualify for AFIT master's degree programs. The specific requirements vary slightly depending on the graduate programs (the specific requirements are listed in AFR 50-5); however, the standards are approximately 400 on the GRE-V, 600 on the GRE-Q, 500 on the GMAT and an undergraduate GPA not less than a 3.0 (AFIT graduate catalog, 1993-1995). These are not rigid requirements; however, their effect on the range of scores is substantial. For example, the average GRE-V score in the general population is

500 with a standard deviation of 100. In our study, the average GRE-V score is 557 with a standard deviation of 80, skewing the distribution . Because reliability measures are intended to be proportional to the ratio of the true-score variance to the observed score variance, the reliability of a truncated sample is distorted (Nunnally: 242). By adjusting for the range attenuation, the overall value of these measures, as related to this study, can be significantly improved.

Correlations

A correlation matrix of the entire data set was constructed to gain insight into the items which have the most potential for predictive significance. Because there are many subjects (overall N=574), most of indicated correlations are significant at the 5% level; however, the uncorrected correlations are not particularly strong. Based on this finding, factor analysis was performed to identify composite factors with increased reliability. The correlations were then adjusted for range attenuation and measurement error as described in Chapter II. The following tables (1 through 4) display the correlations and illustrate the impact of the adjustments.

Definition of Variables

The variables above are defined as follows:

r is the observed or estimated reliability of the measures used.

Δ**r** is a realistic upper bound for reliability which could be achieved by improving the measures.

σ is the overall population standard deviation.

s is the observed standard deviation.

OBJ is the number of multiple choice questions answered correctly on the EPE.

ED is the variable representing Educational items identified through the factor analysis of the student survey. Although it is explained in the following section, it includes items relating to the number of high school and college English courses taken.

ESSAY represents the score received on the essay portion of the EPE.

EXP is the variable representing professional experience items identified through the factor analysis of the student survey. It is also explained in the following section.

GMAT is the total score received on the Graduate Management Admissions Test (GMAT).

UGPA is the student's undergraduate grade point average on a 4 point scale.

GRE-Q AND **GRE-V** are the student's scores on the Quantitative and Verbal portions of the Graduate Record Examination.

WGPA is the measure designed for this study which represents the student's writing grade point average at AFIT. This measure was constructed by averaging the grades each student received in 13 specific courses identified by AFIT faculty members in which a major portion of the grade is based on writing assignments. The average number of these courses taken by an individual student is 2.4 with a standard deviation of 1.5.

GMSA is the construct developed to estimate "Graduate Management School Aptitude." GMSA is a weighted sum of GRE-V and GRE-Q scores. The weights used are based on regressing GRE-V and GRE-Q on GMAT scores. If actual GMAT scores were available, that score was used.

COM310 is a dichotomous variable identifying students who were enrolled in the English refresher course (AFIT identification code of COM 310).

Table 1. Initial Correlation Matrix

ORIGINAL	OBJ	ED	ESSAY	EXP	GMAT	GPA	GRE_Q	GRE_V	WGPA	GMSA	COM310
OBJ	1	0.16	0.14	0.02	0.3	0.14	0.18	0.26	0.14	0.3	-0.38
ED		1	0.14	-0.05	0.11	-0.03	0.11	0.12	0.1	0.11	-0.15
ESSAY			1	-0.01	0.1	0.1	0.01	0.22	0.16	0.1	-0.5
EXP				1	-0.01	0.08	-0.03	0.03	-0.01	-0.01	-0.03

GMAT	1	0.18	0.66	0.51	0.17	1	-0.09
GPA		1	-0.06	0.07	0.12	0.08	-0.07
GRE-Q			1	0.23	0.08	0.85	-0.1
GRE-V				1	0.12	0.58	-0.22
WGPA					1	0.17	-0.16
GMSA						1	-0.12
COM310							1

Table 2. Figures Used for Measurement Error and Range Attenuation Adjustments

	OBJ	ED	ESSAY	EXP	GMAT	UGPA	GRE_Q	GRE_V	WGPA	GMSA	COM310
r	0.62	0.54	0.50	0.70	0.90	1.00	0.90	0.90	1.00	0.90	1.00
Δr	0.80	0.85	0.70	0.85	0.90	1.00	0.90	0.90	1.00	0.90	1.00
σ	4.00	2.67	2.00	2.82	100	0.50	100	100.00	0.50	100	0.42
s	3.25	2.67	1.83	2.82	72.86	0.38	80.76	80.60	0.29	62.67	0.42

Formula 1 was used to convert the initial correlation matrix to the correlation matrix corrected for measurement error.

$$r_{xy_c} = r_{xy}(\Delta r_{xx} * \Delta r_{yy})^{.5} / (r_{xx} * r_{yy})^{.5} \quad (1)$$

Note: The variable choice for X or Y has no impact on the formula, and the Δr values presented in the table above are estimated, realistic goals for reliability which were subjectively chosen based on the reliability achieved in previous research.

Table 3. Correlation Matrix Corrected For Measurement Error

MEASURE	OBJ	ED	ESSAY	EXP	GMAT	GPA	GRE-Q	GRE-V	WGPA	GMSA	COM310
OBJ	1	0.22	0.19	0.03	0.34	0.16	0.2	0.29	0.16	0.34	-0.43
ED		1	0.21	-0.06	0.13	-0.03	0.14	0.15	0.13	0.13	-0.19
ESSAY			1	-0.01	0.11	0.12	0.01	0.26	0.18	0.11	-0.59
EXP				1	-0.01	0.09	-0.04	0.03	-0.01	-0.01	-0.03
GMAT					1	0.18	0.66	0.51	0.17	1	-0.09
GPA						1	-0.06	0.07	0.12	0.08	-0.07
GRE-Q							1	0.23	0.08	0.85	-0.1
GRE-V								1	0.12	0.58	-0.22
WGPA									1	0.17	-0.16
GMSA										1	-0.12
COM310											1

Formula 2 was used to convert the measurement error corrected matrix to the fully corrected correlation matrix.

$$R_{12} = \{r_{12} * S_1/s_1)\}/(1 - r_{12}^2 + r_{12}^2 * S_1^2/ s_1^2)^{.5} \quad (2)$$

Note: In this formula, σ is used for S_1 and denotes the unrestricted population.

Table 4. Fully Corrected Correlation Matrix

corrected	OBJ	ED	ESSAY	EXP	GMAT	GPA	GRE-Q	GRE-V	WGPA	GMSA	COM310
OBJ	1	0.22	0.2	0.03	0.45	0.2	0.25	0.35	0.26	0.5	-0.38
ED		1	0.23	-0.06	0.18	-0.04	0.17	0.19	0.22	0.21	-0.15
ESSAY			1	-0.01	0.16	0.15	0.01	0.32	0.31	0.18	-0.5
EXP				1	-0.01	0.09	-0.04	0.03	-0.01	-0.01	-0.03
GMAT					1	0.23	0.74	0.59	0.28	1	-0.09
GPA						1	-0.07	0.09	0.21	0.13	-0.07
GRE-Q							1	0.28	0.14	0.93	-0.1
GRE-V								1	0.21	0.75	-0.22
WGPA									1	0.21	-0.16
GMSA										1	-0.6
COM310											1

Analysis of Variance

Certain survey items did not lend themselves to proper traditional correlation because they measured unordered qualitative demographic categorical data instead of ordinal data. The two examples of this are the commissioning source and the category of undergraduate degree. The a priori expectation was that Air Force Academy graduates would have better writing skill than graduates from other commissioning sources due the greater career focus and generally higher academic admission standards. Likewise, the expectation was that individuals with non-technical degrees in the Humanities, Social Sciences, and Business would possess better written communication skills than those individuals with technical degrees in Engineering, Mathematics and the other Sciences. This expectation is based on the assumption that courses in the non-technical fields would primarily emphasize written assignments over problem solving and mathematics. To test these hypotheses, an analysis of variance (ANOVA) and the Levene test for

homogeneity of variance were performed for each section comparing the mean objective and essay scores of each group. The following table shows the results of the ANOVA and Levene Tests for the different categories.

Table 5. EPE ANOVA and Levene tests

	ANOVA Fscore	sig	Levene Stat	Sig
OBJ				
Source	.0253	.9750	.9036	.406
Rank	1.4591	.2129	.3635	.835
Degree	.1168	.7326	.6448	.666
ESSAY				
Source	.5264	.5910	.0439	.957
Rank	1.34	.2535	1.3594	.246
Degree	.1168	.7326	3.0990	.079
WGPA				
Source	1.9319	.1459	.4428	.642
Rank	1.4017	.2321	.9497	.435
Degree	.6309	.6763	.4692	.799

The above table indicates there is no significant difference in the mean scores of the different groups; however, there is a significant difference in the variance of the essay scores for students with technical versus non-technical degrees. The variance of the group with technical degrees displayed significantly lower variance.

Multiple Regression

Several multiple regression analyses were completed during this study. For each step, two models were developed, one using the raw data as collected, the other using the adjusted correlations discussed above. For assessing relative significance, the t-significance reported for the raw data should be used because there is no straight forward way to test the significance of the

beta-weights produced using an attenuation-corrected correlation matrix in regression. First, variables GMSA, UGPA, COM310, ED, EXP, OBJ and ESSAY were used to develop a predictive model for WGPA scores. Since the essay and objective tests are used to make the COM 310 assignment, they should appear redundant in a model that includes COM 310. The next set of models (3 and 4) tests the significance of the essay and objective tests without COM 310 in the model. The other variables entered are UGPA and GMSA since they were the only other significant contributors. Models 5 and 6 controlled for the treatment effect of COM 310 by considering only those students who were not enrolled in the written communications review course. Models 7 and 8 constitute the best models developed for predicting writing performance at AFIT. The model diagnostics are presented in Tables 6 - 8.

Table 6. Regression Results- All Students

	Model 1 (Raw Data)		Model 2 (Correlation Data)
Variable	β_i	**t-sig**	β_i
Intercept	2.9812	.0001	2.3293
OBJ	.0018	.6876	.0122
ESSAY	.0131	.0865	.0439
ED	.0063	.1906	.0135
EXP	-.003	.4795	-.0025
GMSA	.0005	.0052	.0006
UGPA	.0773	.0145	.1080
COM310	-.0623	.0879	.0932
Adj r-squared	.05663		.1742
F-value	5.69	.0001	18.294
Total Degrees of Freedom	540		574

Table 7. Regression Results- All Students

	Model 3 (Raw Data)		Model 4 (Correlation Data)
Variable	β_i	**t-sig**	β_i
Intercept	2.9347	.0001	2.7139
OBJ	.0055	.1682	.0131
ESSAY	.0194	.0037	.0392
GMSA	.0005	.0030	.00035
UGPA	.0707	.0228	.1022
Adj r-squared	.0604		.1526
F-value	8.839	.0001	26.845
Total Degrees of Freedom	550		574

In the models, COM310 is a categorical variable coded such that a 0 indicates the student was not enrolled in the English Refresher Course, and a 1 indicates the student was enrolled in the Written Communications Refresher Course. The categorical variable was barely non-significant ($\propto$ = .08 level) in model 1, but the sign on the coefficient became negative for model 2. This suggests there are possible interactions between COM 310 and at least one of the other predictors (i.e., the prediction lines -- one for course takers and one for non-takers -- apparently intersect due to different slopes). Since interaction terms could not be analyzed using the correlation matrix, the raw data was used to analyze the interaction affects utilizing the decision tree (Appendix D). This analysis determined the groups have statistically similar y-axis intercepts and different slopes (t-significance for COM 310 and ED interaction was α=.0096). In other words, the intercept difference caused by COM 310 is very small (and probably unreliable because it is extrapolated back to the y axis, well away from the relevant range of data). A t-test between the groups suggests COM 310 does not fully compensate for the differences between the two groups. The t-test shows a significant difference in the WGPA between the groups (α = .0001). The mean WGPA for students who were enrolled in the course was 3.58 with a standard deviation of .32. The mean WGPA for the other students was 3.69 with a standard deviation of .27. Although this is a statistically "significant" difference, in practical terms both of these grades are in the A- range.

Based on the above table, it appears at least the essay portion of the EPE minimally escapes statistically significance at the $\propto$ = .08 level in model 1. Surprisingly, the coefficient for COM 310 changed signs in model two. The corrected correlations were rechecked to ensure they were entered correctly, but no errors were found. The objective test was found to be an insignificant contributor to the model, and this finding reinforces the need to factor analyze the Objective Test to determine which individual questions or groups of questions are correlated with writing ability (see future research section). The significance of the Essay portion may suggests it adds a small amount of explanatory potential to the models; however, given that an interaction

effect is present, the interpretation of model outputs which contain COM 310 are considered uninterpretable.

The observed interaction effects are between COM 310 and education. Our speculation is that students who have a solid educational foundation in written communications would benefit more from a short review course than those students without the educational foundation. Since the exact effect of COM 310 could not be determined, it was necessary to control for its treatment effect. This was accomplished by splitting the database into two separate groups based on COM 310. The analysis was then repeated using only those students who did not take the written communications review course (roughly three-fourths of the total sample, n = 453 students). A new correlation matrix was produced and adjusted in the same manner as previously discussed and the new matrix was used to perform the regression analysis.

Table 8. Regression Analysis - COM 310 = 0

	Model 5 (Raw Data)		Model 6 (Correlation Data)
Variable	β_i	**t-sig**	β_i
Intercept	2.965	.0001	2.9101
OBJ	.0046	.3468	.0055
ESSAY	.0115	.1634	.0104
ED	.0008	.8719	.0032
EXP	-.0019	.6829	-.001
GMSA	.0005	.0114	.0005
UGPA	.1008	.0337	.0992
Adj r-squared	.0384		.0406
F-value	3.8715	.0009	4.194
Total Degrees of Freedom	426		453

The above data indicate the only predictors which account for significant portions of the sample variation are GMSA and UGPA. The drop in r-squared exhibited between models 2 and 6 warrants some discussion. Again the input data was re-checked and no errors were found. The hypothesis for this difference is that by splitting the data set based on COM 310, the variation in the dependent variable, WGPA, was reduced substantially. This hypothesis is illustrated below. The lines represent the regression lines and the circles represent the sample variability.

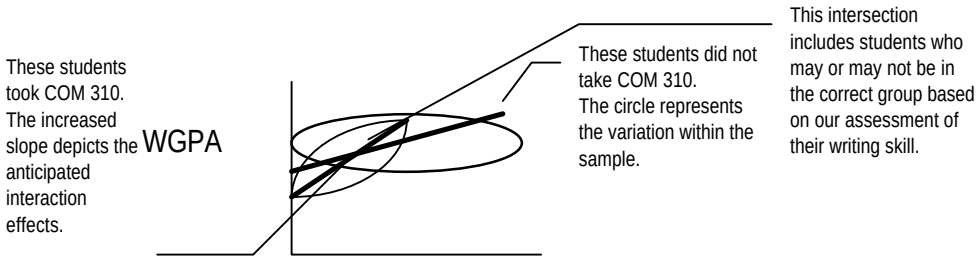


Figure 1. Effect of Sorting Data

The relatively flat circle used to depict the distribution of students who did not take COM 310 indicates little linear variation in WGPA (which accounts for the substantial decrease in r-squared between models 2 and 6). The steeper of the two lines represents the some-what confounded benefits of taking COM 310.

To produce the best possible overall predictive model all models were examined. The goal of the final predictive model was to determine which variables were significant when attempting to estimate the potential students WGPA before they were enrolled in COM 310. The significant variables were ESSAY, GMSA, and UGPA (at the $\alpha = .05$ level)

Table 9. Best Predictive Models Available

	Model 7 (Raw Data)		
Variable	**b_i**	**t-sig**	**β_i**
Intercept	2.982	.0001	0
ESSAY	.01196	.1394	.06979
GMSA	.00057	.0036	.13724
UGPA	.10318	.0018	.14792
Adj r-squared	.04311		
F-value	7.54778	.0001	
Total Degrees of Freedom	433		

The corrected correlation matrix was not used to perform the regression analysis since the predictive model is intended to assist AFIT in the selection process for the written communications review course. Caution should be used when making any judgments as to the relative effects of each of the variables based on significance levels, or beta weights. If comparisions are performed, the standardized beta-weights (β_i) should be used. Given raw data on future students, the following prediction might be cautiously applied:

$$Z_{WGPA} = (.148)Z_{UGPA} + (.137)Z_{GMSA} + (.0698)Z_{ESSAY}$$

or equivalently:

$$WGPA = 2.982 + (.103)UGPA + (.00057)GMSA + (.011967)ESSAY$$

Caution should be used when applying these equations; however the equations are the best possible predictors available.

V. Conclusions and Recommendations for Follow-on Research

Overview

This study was successful in answering the first research question by finding some performance, education, and experience factors which were moderately correlated to writing ability. The predictors which were moderately correlated with our measure of writing ability (WGPA) are as follows (each followed by r-value, significance, N): OBJ (.26, .001, 575), ED(.21, .013, 570), EPE (.25, .000, 573), Essay (.26, .000, 573), GMAT (.28, .008, 246), GRE-V (.21, .019, 375), and UGPA (.21, .004, 567). P-values are based on the uncorrected correlations (Bobko, 1995: 81).

Using correlation adjustments for range attenuation and measurement error, substantially larger correlations are produced. These adjusted correlations were then used to perform a regression analysis in which the predictive utility of the EPE testing procedure (beyond the utility of freely available data) was examined.

Research question two assesses the usefulness of the EPE, and based on regression models 3 and 4, it appears that the essay portion of the EPE is a fairly significant contributor to the regression models ($\propto < .0037$). The insignificance of the objective test ($\propto = .1682$) suggests further investigation is required. There were obviously many factors at work which this study did not account for and therefore manifested themselves as random error in the regression equation. Based on this analysis, the answer to research question 2 is that at least the essay portion of the EPE is worthwhile.

The study was unsuccessful in answering research question three, determining the effectiveness of the written communications review course; however, the decision tree analysis suggests the course may be helpful to those students with a solid foundation in written

communications (i.e., improvement in AFIT writing performance for review course takers seems to depend on prior writing related experience).

Limitations

One of the major limitations of this study is the muddying effect of criterion range attenuation and contamination (e.g., unreliable measures and severe range restriction). Although the variables indicated statistically significant correlation and predictive capability, moderate to weak correlations and the small amount of sample variation accounted for by models make any policy decisions concerning the AFIT writing assessment process problematic. A further limitation of the study is related to from the factors and criteria used in the analysis. Since the objective tests and essays from the EPE have not been saved, it was impossible to gain any insight into reliability of the objective test and the essay scoring protocol. The measure of writing grade point average is subject to a great deal of variability among courses and instructors (i.e., the portion of WGPA variability which systematically reflects stable/enduring differences in writing ability may be quite limited).

As stated in many of the previous studies, writing ability is a difficult construct to measure and even more difficult to predict without a large commitment of time and resources. This study provides more support for that assertion. Requiring applicants to pass new GRE or GMAT writing examinations will ensure the incoming AFIT graduate students are proficient writers.

Recommendations for Follow-on Research

The first recommendation for follow-on research is to collect essay and multiple choice examinations for approximately 400 more students and factor analyze the objective portion and use multiple, trained raters to grade the essays. The construct validity and the reliability of the tests could then be assessed.

Another recommendation for further research is to construct a complete data set from the measurement-error-adjusted correlation matrix. The multiple regression results produced from this simulated data set could be compared to the results produced from using a corrected correlation matrix as the input data file for the regression. The simulated data set could also be compared to a data set of estimated true scores.

Several improvements could be made to the actual measure of writing ability as well. Instead of using only the WGPA from writing intensive courses, writing samples could be gathered from instructors for students throughout their AFIT time (post-remedial training) and an average writing grade could be constructed for each student. This grade would probably more appropriately represent the student's actual AFIT writing performance. A measure of the impact of AFIT on the writing ability of the graduate students could also be estimated via a cross-lagged time series design. Another avenue could be using overall AFIT GPA (non-writing intensive courses) to control for academic performance (and manifest academic dedication) and to estimate the impact of writing ability on AFIT grades as a whole. One would expect that (at the graduate level at least) the ability to write well would have a significant impact on student grades.

Another aspect which could be investigated would be adding trait and state personality variables as predictive factors of effective writing performance. One source of variation which this study failed to address is the amount of effort each student put into the EPE and the writing courses. We might find that conscientiousness, a personality trait meaning thorough or careful, accounts for substantial variation in WGPA not directly attributable to the quasi-experimental treatment (the remedial writing course, COM 310). Controlling simultaneously for conscientiousness and theory-driven ability facets may be a more illuminating approach for evaluating this sort of data.

Appendix A: EPE

DIAGNOSTIC TEST OF WRITTEN SKILLS

Part A. Objective examination of aspects of grammar (30 minutes).

INSTRUCTIONS. Each of the following two sections tests your knowledge of the following common grammar or sentence construction problems: subject-verb agreement, pronoun-antecedent agreement, vague pronoun antecedent, incorrect case form of pronouns, misplaced modifiers, dangling modifiers, non-parallel structure, comma splice, or sentence fragment. You are not asked to name the grammatical errors, only to respond to questions which illustrate those problems. Do not be distracted by the subject matter or style of the sentences--you are NOT ASKED to determine if the sentences are true or capable of stylistic improvement; you are to determine **only if they are grammatically correct**.

Part 1. Correct sentence identification. Indicate whether each of the following statements is a correct or incorrect grammatical sentence, according to the practices of standard written English.

1. Prepared for any emergency, the mission of the unit is to provide air-sea rescue service capability along the Gulf coast.
   a. Correct
   b. Incorrect

2. Those figures are correct, they were revised last week.
   a. Correct
   b. Incorrect

3. The maintenance specialist determined that there were three electrical problems: a faulty warning light, a burned-out fuse, and in one of the circuits a wire was loose.
   a. Correct
   b. Incorrect

4. The reason being that the inconclusive test results were not questioned by the evaluation team members.
   a. Correct
   b. Incorrect

5. That no lost time should result from on-the-job injuries is the goal of the safety office.
   a. Correct
   b. Incorrect

6. The men in the unit looked especially impressive during the parade the reviewing officers commented favorably on their appearance.

a. Correct
b. Incorrect

7. To plan the mission is harder work than flying the mission.
a. Correct
b. Incorrect

8. The shuttle team members were generally happy with the results of their flight, however, they asked to be rescheduled for another mission as soon as possible.
a. Correct
b. Incorrect

9. While waiting for repairs to be completed at the maintenance depot, alternative methods of moving the equipment must be considered.
a. Correct
b. Incorrect

10. The Major inspected every room in the barracks; including the storage closets.
a. Correct
b. Incorrect

Part 2. Correct word or phrase selection. In each of the following sentences, choose the word or phrase that best completes the sentence so that the meaning of the sentence is clear and the sentence is grammatically correct.

11. They gave the responsibility for mission completion to Major Stevens and ____.
a. I
b. me

12. The stewardess asked if any of the passengers would be willing to give up ____ seat(s).
a. his or her
b. their

13. Neither of the candidates ____ acceptable.
a. is
b. are

14. The team expressed ____ support for the coach by applauding when he entered the room.
a. its
b. their

15. The commander will award the prize to ____ completes the clean-up duties first.
a. whoever
b. whomever

16. A platoon of infantry soldiers ____ guarding the entrance to the base.
   a. was
   b. were

17. The instructor asked us to identify each noun or pronoun in the sentence and to describe how ____ functioned in the sentence.
   a. it
   b. they

18. Each of the fourteen aircraft on display _____ inflight refueling capability.
   a. possesses
   b. possess

19. He is one of those briefers who always _____ detailed visual aids.
   a. prepares
   b. prepare

20. The new maintenance officers are Captain Wilson and _____.
   a. he
   b. him

21. Neither the engines nor the fuselage _____ damaged when the aircraft struck the tops of the trees.
   a. was
   b. were

22. The secretaries object to _____ smoking in the office.
   a. our
   b. us

23. Each of the men in the detention unit has been directed to pick up _____ equipment from supply.
   a. his
   b. their

24. Only occasionally _____ there been signs of discord among the faculty members.
   a. has
   b. have

25. _____ shall we ask to be our guest speaker?
   a. Who
   b. Whom

Essay

Part B. Essay (1 hour)

INSTRUCTIONS. Write a short essay (200-300 words) on the topic assigned below. Your essay should be concise and persuasive. Be sure to develop a thesis statement which is supported by specific facts, details, or illustrations. Your essay will be evaluated for

1. good organization,
2. completeness of thought,
3. clarity of style,
4. factual or illustrative support, and
5. correctness of grammar, punctuation, and mechanics.

You will have one hour to complete the essay. Limit your topic for effective discussion in the time available.

Defend or refute this statement:

"The military services should not be involved in humanitarian missions, because peacetime operations differ noticeably from wartime operations; the goal of military training is to wage war, not to make peace."

Limit the topic; draw on your personal experience or professional reading for supporting ideas or examples.

Appendix B: Student Survey

WRITTEN COMMUNICATION SKILLS

STUDENT SURVEY

Department of Graduate Management Systems
Graduate School of Logistics and Acquisition Management
Air Force Institute of Technology

This questionnaire is intended to provide the Graduate School of Logistics and Acquisition Management with a clearer understanding of your education, background, and training in the area of written communications skills. This information will help us to design writing instruction to better suit your needs.

Please fill in your name on the answer sheet and mark the appropriate blocks to match the letters of your name using a number 2 pencil. Record your answers to the questionnaire on the answer sheet also.

STUDENT WRITING PREPARATION QUESTIONNAIRE

1. What is the source of your commission in the armed forces?
   a. ROTC
   b. USAF Academy
   c. OTS
   d. Other
   e. Not applicable (civilian/foreign military)

2. What is your rank?
   a. 2nd Lieutenant
   b. 1st Lieutenant
   c. Captain
   d. Major
   e. Lt Colonel
   f. International Military
   g. Civilian

3. What was the area of concentration of your undergraduate degree?
   a. Humanities
   b. Social science
   c. Business
   d. Engineering
   e. Mathematics/Science
   f. Other

4. How would you describe the value of good writing skills in your career?
   a. Essential
   b. Very important
   c. Moderately important
   d. Somewhat important
   e. Not important

5. How would you assess your high school/pre-college level Written Communications preparation (Written Communications classes)?
   a. Excellent
   b. Good
   c. Average
   d. Fair
   e. Poor

6. What size college or university did you attend?
   a. Very large (over 10,000 students)
   b. Large (5,000-10,000 students)
   c. Medium (2,000-4,999 students)
   d. Small (Under 2,000 students)

7. In how many undergraduate (college/university) writing courses (English) did you enroll?
   a. 0
   b. 1
   c. 2
   d. 3
   e. 4
   f. More than 4

8. In how many undergraduate literature courses did you enroll?
   a. 0
   b. 1
   c. 2
   d. 3
   e. 4
   f. More than 4

9. In how many high school English classes did you enroll?
   a. 0
   b. 1
   c. 2
   d. 3
   e. 4
   f. More than 4

10. How long has it been since you completed your most recent undergraduate (or graduate) degree?

a. 1 year or less
b. Less than 2 years but more than 1 year
c. Less than 3 years but more than 2 years
d. Less than 4 years but more than 3 years
e. Less than 5 years but more than 4 years
f. Less than 6 years but more than 5 years
g. 6 years or more

11. In your most recent assignment, how much of your writing did you do on a computer (word processing)?
a. All
b. Most
c. Some
d. Very little
e. None

12. In general, how would you describe the writing instruction you have received while working for the government or military?
a. Very good
b. More than adequate
c. Adequate
d. Less than adequate
e. Poor

13. Assuming that technical writing includes proposals, mechanism descriptions, process descriptions, or sets of instructions, would you say that you have done some technical writing on the job?
a. Yes, a significant amount
b. Yes, some
c. Yes, a little
d. No

14. Which word processing systems are you familiar with?
a. Word Perfect
b. Word Star
c. Macwrite
d. Other DOS system
e. Other
f. I am not familiar with any word processing system

15. What is your estimation of your writing ability?
a. Well above average
b. Above average

c. Average
d. Below average
e. Well below average

16. Have you ever provided writing instruction to others in your military or civilian profession?
a. Yes, formally (in a class)
b. Yes, informally (working environment -- quite a bit)
c. Yes, informally (working environment -- some)
d. Yes, informally (working environment - a little)
e. No

17. How much has writing or report preparation been a part of your military or civilian career?
a. A crucial component(over 75% of time spent writing)
b. A major component (51-75% of time)
c. A moderate component (26-50 % of time)
d. A minor component (11-25% of time)
e. A negligible component (O-10% of time)

18. How much formal career-related writing instruction have you received?
a. A significant amount
b. A moderate amount
c. A small amount
d. Very little
e. None

19. In what program are you enrolled in the Graduate School of Logistics and Acquisition Management?
a. GAL
b. GCA
c. GCM
d. GIR
e. GLM
f. GSS

20. (Continue question number 19)
a. GMM
b. GSM
c. GTM
d. GIM
e. Environmental Engineering
f. GIS

Appendix C - Factor Analysis Charts

Factor Scree Plot:

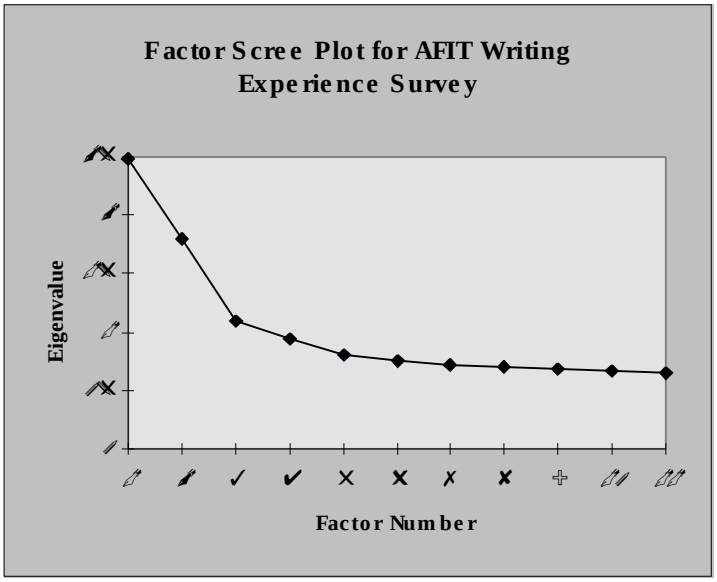


Figure 2. Scree Plot

Reliability Analysis - Scale (A L P H A) - Experience Factor (factor 1)

		Mean	Std Dev	Cases
1.	ITEM_4	1.4709	.6529	1098.0
2.	ITEM_12	3.0273	.9629	1098.0
3.	ITEM_13	2.3151	1.0045	1098.0
4.	ITEM_16	3.3106	1.2328	1098.0
5.	ITEM_17	2.7486	1.0357	1098.0
6.	ITEM_18	3.2532	1.1009	1098.0

Correlation Matrix

	ITEM_4	ITEM_12	ITEM_13	ITEM_16	ITEM_17	ITEM_18
ITEM_4	1.0000					
ITEM_12	.1289	1.0000				
ITEM_13	.1753	.1937	1.0000			
ITEM_16	.2587	.2624	.2610	1.0000		
ITEM_17	.3694	.2116	.3654	.3918	1.0000	
ITEM_18	.1358	.5455	.2352	.3316	.3469	1.0000

Statistics for	Mean	Variance	Std Dev	N of Variables
Scale	16.1257	14.8648	3.8555	6

Item-total Statistics - N of Cases = 1098.0

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Alpha if Item Deleted
ITEM_4	14.6548	12.9354	.3201	.1547	.6953
ITEM_12	13.0984	11.1991	.4249	.3094	.6652
ITEM_13	13.8106	11.3242	.3744	.1620	.6807
ITEM_16	12.8151	9.7589	.4656	.2265	.6547
ITEM_17	13.3770	10.3190	.5220	.3219	.6329
ITEM_18	12.8725	10.0895	.5095	.3707	.6361

Reliability Coefficients 6 items

Alpha = .7022 Standardized item alpha = .7009

Reliability Analysis - Scale (A L P H A) - Education Factor (factor 2)

		Mean	Std Dev	Cases
1.	ITEM_15	3.7783	.7217	1087.0
2.	ITEM_5	4.0690	.7319	1087.0
3.	ITEM_9	4.4471	1.1128	1087.0
4.	ITEM_7	2.8638	1.1083	1087.0
5.	ITEM_8	2.2318	1.1312	1087.0

Correlation Matrix

	ITEM_15	ITEM_5	ITEM_9	ITEM_7	ITEM_8
ITEM_15	1.0000				
ITEM_5	.3497	1.0000			
ITEM_9	.1660	.2255	1.0000		
ITEM_7	.1338	.0740	.1958	1.0000	
ITEM_8	.1634	.0997	.2285	.3256	1.0000

Statistics for	Mean	Variance	Std Dev	N of Variables
Scale	17.3901	8.4462	2.9062	5

Item-total Statistics - N of Cases = 1087.0

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Alpha if Item Deleted
ITEM_15	13.6118	6.8086	.2966	.1470	.4948
ITEM_5	13.3211	6.8885	.2660	.1513	.5075
ITEM_9	12.9430	5.5161	.3237	.1111	.4718
ITEM_7	14.5262	5.5847	.3118	.1259	.4800
ITEM_8	15.1582	5.3433	.3487	.1439	.4542

Reliability Coefficients 5 items

Alpha = .5393 Standardized item alpha = .5496

Pattern and Structure Matrices

OBLIMIN converged in 6 iterations - Kaiser Normalization.

Pattern Matrix

	Factor 1	Factor 2
ITEM_18	.65205	.17874
ITEM_17	.62354	

Structure Matrix

	Factor 1	Factor 2
ITEM_18	.64339	
ITEM_17	.62628	

ITEM_16	.56517	
ITEM_12	.53072	
ITEM_13	.45004	
ITEM_4	.38650	-.20201
ITEM_15		.48439
ITEM_9		.45494
ITEM_5		.45397
ITEM_8		.42954
ITEM_7		.35164

ITEM_16	.57181	
ITEM_12	.52435	
ITEM_13	.44836	
ITEM_4	.39629	-.22074
ITEM_15		.48522
ITEM_5		.46043
ITEM_9		.45090
ITEM_8		.42755
ITEM_7		.35153

Factor Correlation Matrix:

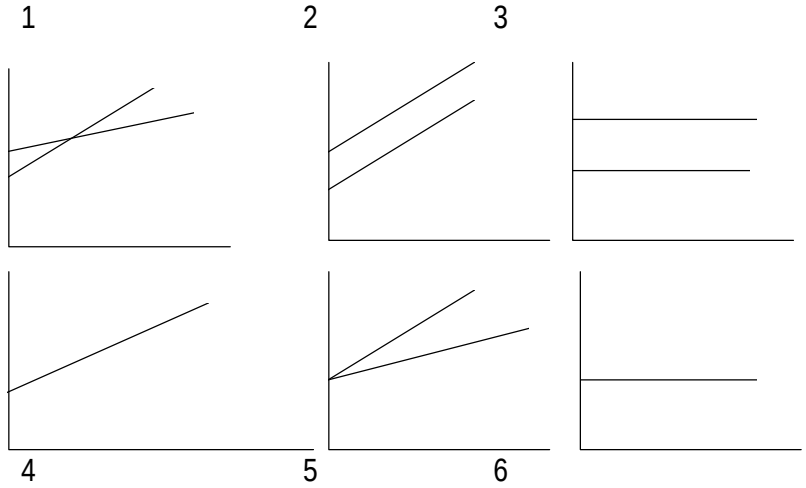
	Factor 1	Factor 2
Factor 1	1.00000	
Factor 2	-.04846	1.00000

Appendix D - Interaction Analysis

Decision Tree for Interaction Effects

The following figures depict six different models used for determining the affect categorical variables their interaction terms on linear regression models. The ($c_i D_i$) terms are the dummy variables and the ($d_i D_i X$) terms are the interaction terms. Test for significance using the formula below and follow the decision tree below to determine which situation applies to the specific situation.

$$F_{\Delta r}^{\ 2} = \frac{(R_1^{\ 2} - R_2^{\ 2})/(k_1 - k_2)}{(1 - R_1^{\ 2})/(N - k_1 - 1)}$$



Model	Formula	Model	Formula
1.	$Y' = a + bx + \sum c_i D_i + \sum d_i D_i X$	4.	$Y' = a + bx$
2.	$Y' = a + bx + \sum c_i D_i$	5.	$Y' = a + bx + \sum d_i D_i X$
3.	$Y' = a + \sum c_i D_i$	6.	$Y' = a$

Figure 3. Interaction Illustration

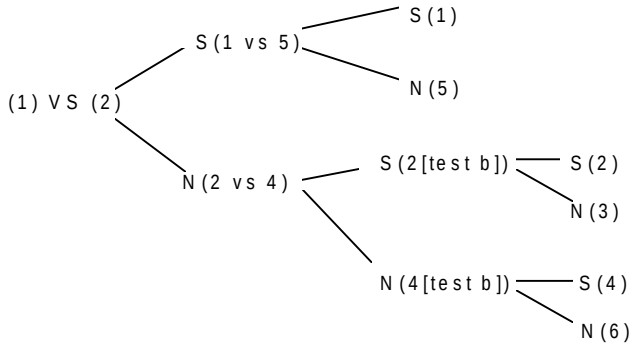


Figure 4. Decision Tree

To use this decision tree, compare the models in parenthesis for significant differences using the F formula above. If the F value is significant, follow the “S” branch, if not, follow the “N” branch.

References

Bobko, Philip Correlation and Regression: Principles and Applications for Industrial/Organizational Psychology and Management New York: McGraw-Hill Inc., 1995.

Bradley-Johnson, Sharon, and Judi Lesiak. *Assessment of Written Expression: A Critique of Procedures and Instruments*. Brandon VT: Clinical Psychology Publishing, 1981.

Breland, Hunter M., Roberta Camp, Robert J. Jones, Margaret M. Morris, and Donald A. Rock. *Assessing Writing Skill*. Research Monograph No. 11. New York: College Entrance Examination Board, 1987.

Breland, Hunter M. The Direct Assessment of Writing Skill: A Measurement Review. College Board Report No. 83-6. New York: College Board Publications, 1980.

Department of the Air Force. AFIT Graduate Catalog 1993 - 1995. WPAFB: AFIT, 1993.

Diederich, P. *Measuring Growth in English*. Urbana IL: National Council of Teachers of English, 1974.

Educational Testing Service. Interpreting GMAT Results. Graduate Management Admission Council. Princeton: Educational Testing Service, 1996.

Follman, J., and J. Anderson. "An Investigation of the Reliability of Five Procedures for Grading English Themes." *Research in Teaching of English: 1*:190-200.

Freedman, Sarah Warshauer, Anne Haas Dyson, Linda Flower, and Wallace Chafe. *Research in Writing: Past, Present, and Future*. Technical Report No. 1. Berkeley CA and Pittsburgh PA: Center for the Study of Writing, 1987.

Huddleston, E. "Measurement of Writing Ability at the College Entrance Level: Objective vs. Subjective Testing Techniques." Journal of Experimental Psychology 22: 165-213. (1954).

Nilsson, Johanna E. "The GRE and GMAT: A Comparison of Their Correlations to GGPA." Educational and Psychological Measurement, 55: 637-640 (August 1995).

Norman, Cliff. Analyzing Multivariate Data, New York: Academic Press, 1987.

Nunnally, Jum C. Psychometric Theory, New York: McGraw-Hill Publishing Company, 1978.

Pedhazur, Elazar J. Multiple Regression In Behavioral Research: Explanation and Prediction New York: The Dryden Press, 1982.

Schmidt, Frank L. and John E. Hunter. "Measurement Error in Psychological Research: Lessons From 26 Research Scenarios," Psychological Methods, 1: 199-223 (1996).

Sheppard, E. "The Effect of Quality of Penmanship on Grades." Journal of Educational Research 19: 102-5.

Spandel, Vicki and Richard J. Stiggins. Direct Measures of Writing Skill: Issues and Applications Portland: Clearinghouse for Applied Performance Testing, 1980.

Van Scotter, James R. Predictors of Success in Air Force Institute of Technology Resident Master's Degree Programs A Validity Study. MS Thesis, AFIT/GLM/LSSR/83S-4. School of Systems and Logistics, Air Force Institute of Technology (AU), Wright Patterson AFB OH, September 1983 (AD-A134391).

Vita

Captain Darrin E. Farr earned his Bachelor of Science Degree in Engineering Physics from Miami University, Oxford, OH, in 1991. Upon graduation, he was assigned as an Operations Analyst at the 414th Combat Training Squadron (Red Flag), Nellis AFB NV. In 1994, he was assigned to the 554th Range Squadron as the Assistant Program Manager for New Acquisitions. Upon graduation from the AFIT Graduate Cost Analysis Program in September 1996, Capt Farr will be assigned to the Aeronautical Systems Center Reconnaissance SPO, Wright-Patterson AFB, OH.

Permanent Address:

2 Crescent Ave
Warwick, NY 10990

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1996	Master's Thesis

4. TITLE AND SUBTITLE
AN EVALUATION OF THE FACTORS USED TO PREDICT WRITING ABILITYAT THE AIR FORCE INSTITUTE OF TECHNOLOGY

5. FUNDING NUMBERS

6. AUTHOR(S)
Darrin E. Farr, Captain, USAF

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)

Air Force Institute of Technology
2950 P Street
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT/GCA/LAS/96S-6

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Educational Testing Service
Princeton, NJ 08541-6106

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 Words*)

A study of 574 students at the Air Force Institute of Technology compared performance, education, and experience factors--the later two as stated by the students themselves--to a locally developed estimate of true writing ability (WGPA). This exploratory research was additionally intended to assess the effectiveness of AFIT's current writing student skill diagnostic and instructional system. Direct (essay evaluation) and indirect (objective test) evaluations of AFIT student writing ability were analyzed for their predictive impact. The statistical analysis procedures used in this study included the factor analysis of a survey, ANOVA, the adjustment of multiple correlations due to measurement error and range attenuation, and the performance of a regression analysis using the raw data and the adjusted correlation matrix. The results of this study indicate AFIT's direct evaluation portion (essay examination) is useful for determining writing ability; the indirect portion (objective test) did not significantly contribute to the model. Due to the combination of independent variables chosen for the predictive model, the study was unable to identify the immediate benefits of the written communications review course on AFIT performance.

			15. NUMBER OF PAGES 74
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED

Abstract

The focus of this research is to determine whether performance, education, and experience factors can be correlated with writing ability, and to conduct an initial assessment of the current Air Force Institute of Technology (AFIT) written communications review course. A study of 574 students at the Air Force Institute of Technology compared performance, education, and experience factors--as stated by the students themselves--to a locally developed estimate of true writing ability (WGPA). This exploratory research was additionally intended to assess the effectiveness of AFIT's current writing student skill diagnostic and instructional system. Direct (essay evaluation) and indirect (Objective Test) evaluations of AFIT student writing ability were analyzed for their predictive impact.

The statistical analysis procedures used in this study included the factor analysis of a survey, ANOVA, the adjustment of multiple correlations due to measurement error and range attenuation, and the performance of a regression analysis using the raw data and the adjusted correlation matrix.

The results of this study indicate AFIT's direct evaluation portion (essay examination) is useful for determining writing ability; however, due to the combination of independent variables chosen for the predictive model, the indirect portion (objective test) did not significantly contribute to the model. The study was also unable to identify the immediate benefits of the written communications review course on AFIT performance.